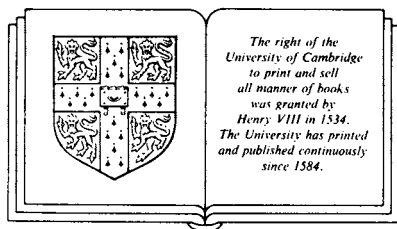


Speech physiology, speech perception, and acoustic phonetics

Philip Lieberman and Sheila E. Blumstein

Department of Cognitive and Linguistic Sciences, Brown University



Cambridge University Press

Cambridge

New York New Rochelle Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1988

First published 1988

British Library cataloguing in publication data

Lieberman, Philip

Speech physiology, speech perception, and
acoustic phonetics. – (Cambridge studies in
speech science and communication).

I. Speech

I. Title II. Blumstein, Sheila

612'.78 QP306

Library of Congress cataloguing in publication data

Lieberman, Philip.

Speech physiology, speech perception, and acoustic phonetics.
(Cambridge studies in speech science and communication)

Bibliography.

Includes index.

I. Speech – Physiological aspects. 2. Speech
perception. 3. Phonetics, Acoustic. I. Blumstein,
Sheila. II. Title. III. Series.

QP306.L53 1988 612'.78 87–13187

ISBN 0 521 30866 6 hard covers

ISBN 0 521 31357 0 paperback

Transferred to digital printing 2002

Contents

<i>List of figures</i>	ix
<i>Preface</i>	xiii
<i>Acknowledgements</i>	xv
1 Introduction	1
2 A qualitative introduction to the physiology of speech	3
The three physiological components of speech production	3
The subglottal respiratory system	5
The larynx	10
The supralaryngeal vocal tract	13
The neural processing of speech	14
3 Basic acoustics	16
Graphs and physical measurements	16
Sinusoidal waves	20
Fourier analysis	24
Amplitude and frequency scales for speech	27
Filters	29
4 Source-filter theory of speech production	34
The laryngeal source	34
The supralaryngeal filter	36
The perception of fundamental and formant frequencies	38
Formant frequency calculation	42
Formant lowering and vocal tract length	47
5 Speech analysis	51
The sound spectrograph	52
Interpreting spectrograms – how the spectrograph works	59
Measuring formant frequencies on spectrograms	64
Tape recording techniques	73
Computer-implemented procedures	77
6 Anatomy and physiology of speech production	90
The lungs	91

Contents

The larynx	97
The supralaryngeal vocal tract	114
7 Speech synthesis and speech perception	140
Vocoder synthesizers	140
Speech synthesis and segmentation	142
Speech encoding	145
The “motor theory” of speech perception	147
The speech “mode” of perception	148
Neural acoustic property detectors	149
Electrophysiological and comparative studies	150
Psychoacoustic tests	152
Critical bands	159
8 Phonetic theories	162
Traditional “articulatory” phonetic theory	163
Vowels	164
Phonetic features	183
Prosodic features	198
Linguistic universals and biological structuring	203
9 Some current topics in speech research	205
The evolution of human speech	205
Ontogenetic development of speech	209
Speech pathologies	213
Aphasia	214
10 Acoustic correlates of speech sounds	221
Vowels	221
Stop consonants	224
Nasal consonants	226
Liquids and glides	226
Fricatives	227
<i>Bibliography</i>	228
<i>Index</i>	242

Figures

2.1 The three physiologic components of human speech production.	4
2.2 A mechanical model of the subglottal respiratory system.	6
2.3 Some of the muscles that control respiration.	9
2.4 Three frames from a high-speed movie showing the vocal cords gradually moving together before the start of phonation.	12
2.5 Three frames from the same high-speed movie as in Figure 2.4, exposed during phonation.	12
3.1 Graph of temperatures plotted from data in Table 3.1.	18
3.2 Graph of water height showing the 10 second period of the waveform.	19
3.3 Sinusoidal wave, with the same frequency but half the amplitude of that in Figure 3.2.	21
3.4 Wave of Figure 3.2 shifted in phase.	22
3.5 Propagation of an ocean wave.	23
3.5 A complex waveform and its Fourier components.	25
3.7 Graph of a spectrum.	27
3.8 Sieve as mechanical filter.	29
3.9 Transfer function of sieve as “small-diameter pass through” filter.	30
3.10 Transfer function of sieve as “large-diameter pass through” filter.	31
3.11 Transfer function of two sieves used to make a “band pass” filter.	32
4.1 Three cycles of a waveform typical of the air flow through the glottis during phonation.	35
4.2 Spectrum of typical glottal air flow.	35
4.3 Transfer function of the supralaryngeal vocal tract for the vowel [ə].	37
4.4 The spectrum that would result if the transfer function plotted in Figure 4.3 were “excited” by the glottal source plotted in Figure 4.2.	38
4.5 Speech waveform as measured by a microphone sensitive to air pressure variations.	39
4.6 Spectrum of the sound [ə] produced with a different fundamental frequency of phonation.	40
4.7 (A) Spectrum of glottal source. (B) Transfer function for vowel [i]. (C) Spectrum of the speech signal that would result if transfer function (B) were excited by glottal source (A).	41
4.8 The adult human vocal tract.	43
4.9 A uniform tube open at one end, showing pressure waveform at the first formant frequency.	44
4.10 The relation of quarter-wavelength “inside” the tube to the air pressure wave waveform outside the tube.	44
4.11 The uniform tube of Figure 4.9 showing the second formant frequency, which is three times higher than the first formant frequency.	47
4.12 Formant frequency lowering that results from a constriction at the end of a uniform tube.	48

List of figures

5.1 A plot of signal strength of FM radio stations.	55
5.2 Wide-band spectrogram of the sound [ba].	56
5.3 Narrow-band spectrogram of the same sound [ba] analyzed in Figure 5.2.	58
5.4 Sketch of a hypothetical "photo-output" narrow-band sound spectrograph machine.	60
5.5 Spectrum of a speech signal and two stages of processing in the spectrograph machine.	62
5.6 The effect of "mark level" on a spectrogram.	63
5.7 Functional operation of wide-band filtering of spectrograph.	65
5.8 Operation of wide-band filter when a formant frequency does not coincide with that of a harmonic of the fundamental frequency.	66
5.9 The spectrographic analysis of a speaker whose fundamental frequency exceeds 200 Hz.	68
5.10 The "staircase" effect on the wide-band spectrogram.	70
5.11 Spectrogram of breathy utterance of a newborn infant.	71
5.12 Using a narrow-band, 50 Hz bandwidth spectrogram to estimate formant frequencies.	72
5.13 Using the mirror image technique for "zero" frequency calibration.	73
5.14 The waveform of the sound [ba] as it is displayed on the monitor of the BLISS system.	80
5.15 The waveform of a [ta] with cursor L0 representing the onset of the burst and R0 representing the start of phonation.	81
5.16 Waveforms for the sounds [ba] and [wa].	81
5.17 Waveforms and fundamental frequency of phonation for the vowel [i] in two words of Chinese spoken in isolation.	83
5.18 A waveform with successive pitch periods marked by the computer.	84
5.19 A plot of the fundamental frequency of the word <i>dog</i> spoken by a three-month-old child.	85
5.20 A speech waveform for the vowel [i] with a full-Hamming window placed after the third pitch period.	85
5.21 Discrete Fourier transform of the vowel [i] shown in Figure 5.20.	86
5.22 Formant frequencies derived for the speech sample under the window shown in Figure 5.20.	87
5.23 A three-dimensional display of the LPC spectrum for the vowel [i] shown in Figure 5.20.	88
6.1 (A) Pulmonary air pressure during quiet respiration. (B) Pulmonary air pressure during speech.	91
6.2 The electrical activity of several muscles of the chest during speech.	92
6.3 "Feedback" system for regulation of a respiratory muscle.	95
6.4 Cartilages and intrinsic muscles of the larynx.	98
6.5 Action of the intrinsic laryngeal muscles.	99
6.6 Schematic view of vocal cords and some of the forces of phonation.	102
6.7 Subglottal air pressure, acoustic signal, and electromyographic activity of cricothyroid and sternohyoid muscles.	105
6.8 Computer-averaged and computer-integrated electromyographic signals and fundamental frequency of phonation as functions of time for 38 tokens of the sentence <i>Bev loves Bob</i> spoken by one speaker.	106
6.9 Plots similar to those of Figure 6.8 for the same speaker's tokens of the question <i>Bev loves Bob?</i>	107
6.10 Adjustments of the larynx that can produce normal (A) and breathy phonation (B and C).	111

6.11 Measurement of glottal area waveform derived from computer analysis of a high-speed movie.	112
6.12 Stylized midsagittal section of adult human vocal tract.	115
6.13 The pharyngeal constrictor muscles.	116
6.14 The “strap” muscles that connect the mandible, skull, larynx, and breastbone.	117
6.15 Some of the muscles of the tongue.	118
6.16 Muscles of the face and jaw.	119
6.17 Diagram of a frame from an X-ray movie of the vocal tract of an adult human male during the production of speech.	125
6.18 Midline contours of the tongue for the same speaker as shown in Figure 6.17.	126
6.19 A mechanical model for production of fricative consonants.	133
6.20 Sketch of water supply system.	134
6.21 Electrical analog of water supply system.	134
6.22 Controllable electrical vocal tract model.	137
6.23 Vocal tract parameters controlled by Stevens and House (1955).	138
7.1 Vocoder block diagram.	141
7.2 (A) Sound spectrogram. (B) Simplified version of the same phrase that serves as input to pattern playback.	142
7.3 Synthetic spectrograms using only F_1 and F_2 information that produce the voiced stops before various vowels.	144
7.4 Two formant patterns that will produce the sounds [di] and [du].	146
7.5 The Mel scale which relates perceived pitch to frequency.	155
7.6 Identification and discrimination functions for three different listeners responding to synthetic speech stimuli that are cued by the second formant transitions schematized in Figure 7.7.	156
7.7 Schematic representation of two formant frequency patterns that produce the sounds [bæ], [dæ], and [gæ].	157
7.8 Critical bandwidth as a function of the frequency at the center of the band.	160
8.1 The IPA vowel “quadrilateral.”	164
8.2 Radiographic data derived from an X-ray movie of a speaker producing various English vowels.	166
8.3 Radiographic data for a second speaker.	167
8.4 Radiographic data for a third speaker, whose tongue contours are in accord with the traditional articulatory classification.	168
8.5 (A) Two-tube model of the supralaryngeal vocal tract for the vowel [a]. (B) Area function for two-tube model with air pressure waveforms sketched in.	172
8.6 The first formant frequency F_1 and the second formant frequency F_2 for the two-tube model of Figure 8.5.	175
8.7 Midsagittal sections, cross-sectional area functions, and acoustic transfer functions of the vocal tract for the vowels [i], [a], and [u].	176
8.8 Classifying vowel sounds in terms of their first and second formant frequencies.	177
8.9 Average values of formant frequencies derived by Peterson and Barney for adult male speakers, adult female speakers, and adolescents.	178
8.10 Acoustic classification of the vowels of Swedish in terms of Mel values of first formant and “equivalent” second formant frequency.	180
8.11 Acoustic classification of the vowels of American English.	182
8.12 An idealized model of the supralaryngeal vocal tract for the production of fricative consonants.	185

List of figures

8.13 (A) Results of a computer modeling of the model of Figure 8.12. (B) Sketches of supralaryngeal vocal tract for labial, dental and velar stops.	187
8.14 Examples of waveforms and spectra at the release of three voiced stop consonants as indicated.	189
8.15 Schematized spectrographic representation of a stimulus with formant trajectories appropriate for the syllable [da].	191
8.16 Three-dimensional frequency, amplitude and time display of [bi] and [wi].	194
8.17 Fundamental frequency contours derived from one speaker's production of five tokens of the sentence <i>Bev loves Bob</i> .	201
9.1 Midsagittal view of the head of an adult chimpanzee.	206
9.2 Midsagittal view of the adult human supralaryngeal vocal tract.	207
9.3 Gradual development of the formant frequency pattern for the vowels of American English in an infant from age 16 to 62 weeks.	212
9.4 Computer modeling of vowel production in an Apert's subject.	214
9.5 Waveform of a portion of a syllable beginning with a voiced and voiceless stop consonant.	216
9.6 The distribution of VOT productions for the alveolar stop consonants for a normal control, a Broca's aphasic, and a Wernicke's aphasic.	217
9.7 Discrimination of [ba da ga] stimuli for normals and aphasics.	220
10.1 Mean values of formant frequencies for adult males of vowels of American English.	222
10.2 Formant transitions and bursts for the syllables [ka da ga pa ta ka].	224

Introduction

The study of language and the sounds of speech can be traced back at least to the Greek and Sanskrit grammarians of the third and fourth centuries BC. The explicit study of speech science began in the eighteenth century when Ferrein (1741) attempted to explain how the vocal cords produced phonation. Ferrein's studies were not an isolated event. Kratzenstein (1780) and von Kempelen (1791) attempted to explain how the vowels and consonants of human speech were produced by synthesizing sounds using artificial "talking machines." There indeed may have been earlier attempts at constructing talking machines; La Mettrie (1747) discusses some of these early attempts, but we lack detailed records. By the mid nineteenth century Müller (1848) had formulated the source-filter theory of speech production, which is consistent with the most recent data and still guides research on human speech as well as the vocal communications of other animals. Although Müller's theory was further developed later in the nineteenth century, particularly by Hermann (1894), the modern period of speech science is really quite recent, dating back to the late 1930s, where devices like the sound spectrograph, and techniques like high-speed photography, cineradiography, and electromyography made new data available. Quantitative studies like those of Chiba and Kajiyama (1941), Joos (1948), Peterson and Barney (1952), Stevens and House (1955), and Fant (1960) refined and tested the traditional phonetic theories of the nineteenth century and provided the framework for comprehensive, biologically-oriented studies of speech production, speech perception, and phonetics. The advent and general availability of digital computers made quantitative modeling studies possible. New techniques for speech synthesis and psychoacoustic experiments have made it possible to explore the fundamental properties of human speech.

We are beginning to understand how human speech is produced, how it is perceived, and how the physiological properties of the vocal tract and the neural mechanisms of the brain contribute to speech processing. We also are beginning to understand how human language and human speech evolved and how other animals communicate. The development of speech and language in infants and children is being explored, and new possibilities are opening for the

Introduction

diagnosis and amelioration of various speech pathologies.

The focus of this introduction to speech physiology and acoustic phonetics is thus to provide a background to the “new” speech science. An understanding of the acoustics of speech, the physiology of speech production, and the special factors that are involved in the perception of speech is a prerequisite for further study of the pathologies of speech production or the neurological impairment of either speech production or speech perception. It is also necessary for the development of quantitative, predictive phonetic and phonological studies. While linguists have studied the sound structure of language, exploring the processes of sound change and the structure of sound systems in language, it is the study of speech science which may provide the explanations for *why* sounds change in the way they do and *why* the sound systems of natural language are structured in the way that they are. This introduction is no substitute for a traditional phonetics text, focused on teaching people how to make transcriptions of various languages and dialects. The training techniques that phoneticians use are not included in this book because our objective is to understand the biological mechanisms that are the basis not only of human speech, but also of vocal communication in many other animals.

Readers who have a good background in high-school mathematics should have little difficulty in following the discussions of the acoustics of speech production or the source-filter theory of speech production. Readers who have a more advanced background may be able to skim appropriate chapters.

Although readers may find this book a useful reference source, its primary function is pedagogic. It should be viewed as an introduction to the physiology of speech and acoustic phonetics. Many current problems are not discussed in detail, and the advanced reader may be familiar with topics that have been omitted. Everyone, however, should encounter new material and indeed should note that speech science is still a frontier with many gaps in our knowledge yet to be filled. It is becoming apparent that human speech is an integral part of human linguistic ability. The biological bases of human speech are complex and appear to follow from the Darwinian process of natural selection acting over at least 250 000 years to yield specialized anatomical and neural mechanisms in *Homo sapiens* (Lieberman, 1984). The gaps in our knowledge concerning the biological bases of human speech thus reflect the difficulties inherent in understanding the nature of language and human cognition of which speech forms a crucial part.